DOCUMENT RESUME

ED 157 925 TM 007 283

AUTHOR Jensen, Marjane; Beck, Michael D.
TITLE Gender Balance Analysis of the Metropolitan Achievement Tests, 1978 Edition.
PUB DATE Mar 78
NOTE 19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Toronto, Ontario, Canada, March, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Tests; *Content Analysis; Elementary Education; Evaluation Criteria; *Sex Fairness; Sex Role; *Sex Stereotypes; *Test Bias; *Test Reviews
IDENTIFIERS *Metropolitan Achievement Tests 1978

ABSTRACT

The 1978 edition of the Metropolitan Achievement Tests was analyzed for sex stereotyping and for the use of male, female, or neuter references and a comparison was made with the 1970 edition. There was less bias in the new edition, and there was relatively little sex stereotyping with respect to occupations, activities, and roles for females. There was, however some sex stereotyping with respect to the display of emotions. The index of sex bias used was the ratio of male nouns and pronouns to female, with a correction for repeated references to the same person. A similar index computed the ratio of neutral personal nouns and pronouns to gender-specific words. Illustrations were coded similarly for gender ratios and also analyzed for the comparative stature of males and females. The possible effects of sex bias and sex stereotyping in test content on test performance are yet to be determined; test developers have assumed that an even balance of male and female references would minimize test bias. The 1978 edition of the MAT is more balanced than the 1970 edition, and generally presents a less stereotyped portrayal of both sexes. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

GENDER BALANCE ANALYSIS OF THE METROPOLITAN ACHIEVEMENT TESTS, 1978 EDITION

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Marjane Jensen
Indiana University

Michael D. Beck
The Psychological Corporation

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Marjane Jensen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED157925

Recent years have brought increased attention to gender balance in evaluation devices. Research indicates that current educational materials present males significantly more often than females, and both sexes almost exclusively in stereotypic roles. In revising the Metropolitan Achievement Tests, the publisher attempted to balance the presentation of sexes and to portray both sexes in less stereotypic roles. This study was undertaken to determine whether the revision conformed to these guidelines. Results indicate the test content is significantly better balanced and less stereotypic than previous tests. Breakdowns of test content and various thematics categories are presented; data are compared across content areas and test levels.

TM007 283

Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto, March 1978.

"Printed in U.S.A."

GENDER BALANCE ANALYSIS OF THE METROPOLITAN ACHIEVEMENT TESTS, 1978 EDITION

Beginning with a NOW publication regarding the sex-role stereotyping in public school reading texts (Committee on Women in Words and Images, 1972), there has been a growing body of literature concerning such stereotyping in all curriculum materials, including standardized tests. Several studies and position papers have been published in connection with the issue (c.f., Beck, 1975; Educational Testing Service, 1976; Faggen-Steckler, McCarthy, & Tittle, 1974; Saario, Jacklin, & Tittle, 1973).

Faggen-Steckler, et al. (1974) looked specifically at the item content of eight standardized achievement tests. Their purpose was to examine "sex imbalance as well as... stereotyped representations of women." Their measure of sex imbalance was the ratio of male nouns and pronouns to female nouns and pronouns. A balanced test would have a ratio of approximately 1.0; female nouns and pronouns should be used in items as often as male nouns and pronouns. The results showed ratios ranging from .86 (100 female references for every 86 male references) to 14.0 (14 male references for each female reference).

In his guidelines for reducing "sex bias," Beck (1975) suggested that males and females should be represented with equal frequency within a full range of activities and vocations. Moreover, attempts should be made to maximize the use of neuter references in standardized tests. This could be accomplished by use of neuter titles or generic descriptions, or by re-wording gender-specific, singular phrases to form neutral, plural phrases.

Educational Testing Service (1976) and the McGraw-Hill Book Company (1974), suggest similar guidelines for their publications. ETS adds the warning that re-phrasing should not result in wording that is complex or confusing.

Current commercial achievement tests are not sufficiently balanced in regard to sex-role references (Faggen-Steckler, et al., 1974). Therefore, as new tests are published and existing tests are revised, there will probably be increasing attention paid to sexist terminology and overtly biased attitudes. The present analysis was undertaken to determine the extent to which the 1978 edition of the Metropolitan Achievement Tests (MAT) conforms to certain guidelines for attaining a balanced representation of sexes and sex roles. The test was examined with respect to its overall gender-balance (i.e., use of gender-specific nouns and pronouns), as well as its portrayal of males and females in diverse activities.

In the final version, the 1978 MAT consists of two forms (J and K) of each of four components: Instructional Tests in Mathematics, Reading, and Language and a Survey Battery. The Instructional Tests are available at six levels (Primer to Advanced 1) and the Survey is available at eight levels (Preprimer to Advanced 2). The present investigation includes only Form J of all components and levels.[1]

Procedure

All test booklets (text and illustrations) of the 1978 MAT were analyzed. The test content was examined using both a quantitative and a qualitative method.

---

[1]The authors are grateful to The Psychological Corporation for allowing them access to the pre-publication materials.

The quantitive method was a variation of the one used by Faggen-Steckler, et al., (1974). A tally of male nouns and pronouns, female nouns and pronouns, and neuter nouns and pronouns was made across all components and levels.

Two ratios were calculated:

1) The Gender Ratio (GR)

$$\frac{\text{Male Nouns (MN)} + \text{Male Pronouns (MP)}}{\text{Female Nouns (FN)} + \text{Female Pronouns (FP)}}$$

2) The Neuter Ratio (NR)

$$\frac{\text{Neuter Nouns (NN)} + \text{Neuter Pronouns (NP)}}{(\text{MN} + \text{MP}) + (\text{FN} + \text{FP})}$$

The MAT authors attempted to reduce use of gender-specific terms in favor of gender-neutral terms such as "children", "people", "they", "we", etc. It was thought that the second ratio might reflect the degree of success in this effort to neutralize the language.

The qualitative analysis involved coding the test items according to four broad categories: active or passive roles, and traditional and non-traditional occupations, activities, and emotions (See Appendix A). Items were either coded in one or more category or judged to be neutral. Those judged neutral were assigned to none of the preceding categories.

The following guidelines were used for coding the written material:

1) A noun or pronoun was considered to be neuter when it was used in place of a female or male reference. For example, "people" (instead of "men") was considered a neuter noun; however references such as "the dog...it" were not considered neuter, but were omitted from the counts.

2) Each noun or pronoun was counted the first time it occured in a passage or an item. Repeated occurrences within the same passage or in items referring to the passage were not counted.

3) The thematic codes (occupations, activities, roles, and emotions) were applied only to the correct responses for each item. Nouns and pronouns were counted in correct and distractor responses.

4) Boys and girls engaged in school-related activities were judged to be in male or female traditional activities.

5) References to "mother," "father," "grandmother," or "grandfather" were coded as occurrences of traditional activities for each gender.

All illustrations were coded along the same thematic dimensions. Frequency counts of males and females were made and categorized with respect to traditional or non-traditional clothing, and their physical stature (e.g., male and female figures the same height, male or female taller). See Appendix A. A tally was also kept of children and adults in mixed-sex or same-sex groups.

The following guidelines were applied to the coding of illustrations:

1) Only the correct responses were analyzed along the thematic dimension. All responses were included in the frequency counts of males and females, types of clothing, physical stature and group composition.

2) Physical stature was analyzed within age groups: adults compared with adults, children with children. However, if it was clear one child was drawn larger than another to indicate a difference in age, this was not coded as a difference in physical stature.

Results and Discussion

Table 1 shows the Gender Ratio (GR) for each level within each component of the 1978 MAT. The GR's range from .50 to 2.04 with an overall median GR of 1.02. The Instructional Mathematics Tests, with a median GR of .86, displays the greatest disparity from the optimal GR of 1.00; and the Survey Batteries (median GR of 1.11) show the greatest disproportion of males. As a whole, however, Form J of the MAT seems to have a good gender balance.

In Table 2, GR's from the 1978 MAT Survey Batteries are compared with the data presented by Faggen-Steckler, et al., (1974) for the 1970 edition of MAT. Faggen-Steckler, et al. computed a ratio of male "regular" and "generic" nouns and pronouns to female "regular" and "generic" nouns and pronouns. They defined generic nouns and pronouns as those which "refer to males or females or to a combined group" and for which "there is only one way for referencing the noun in question" (p. 153). The present analysis classified the generic terms as either male, female, or neuter based upon their pronoun reference. It appears that both methods would give essentially the same result with respect to whether a noun or pronoun was male or female. Therefore, it was concluded, that the GR's for the 1978 MAT could be compared with the data of Faggen-Steckler, et al. (1974) for the 1970 MAT. Data for comparable 1970 and 1978 levels only are shown.

The 1970 MAT had a range of ratios from .86 to 4.11 with a median of 2.95. Clearly, the 1978 MAT survey, with a range of .87 to 1.91 and a median of 1.21 shows marked improvement over the earlier edition.

TABLE 1

Gender Ratios (GR) of Male Nouns & Pronouns to Female Nouns & Pronouns by MAT Component and Level

Instructional Test	MN + MP / FN + FP	GR	Median GR by level for Component
Mathematics			
Primer Level	13/16	.81	
Primary 1	17/21	.81	
Primary 2	21/24	.88	.81
Elementary	26/20	1.30	
Intermediate	9/15	.60	
Advanced 1	6/12	.50	
Total: Mathematics	92/108	.85	
Reading			
Primer	28/16	1.75	
Primary 1	56/59	.95	
Primary 2	46/49	.94	1.06
Elementary	47/44	1.07	
Intermediate	45/43	1.05	
Advanced 1	67/33	2.03	
Total: Reading	289/244	1.18	
Language			
Primer	30/38	.79	
Primary 1	53/36	1.47	
Primary 2	64/62	1.03	1.06
Elementary	75/74	1.01	
Intermediate	55/51	1.08	
Advanced 1	50/32	1.56	
Total: Language	327/293	1.12	
Survey Batteries			
Preprimer	22/33	.67	
Primer	31/31	1.00	
Primary 1	74/61	1.21	
Primary 2	66/76	.87	1.11
Elementary	69/68	1.01	
Intermediate	63/33	1.91	
Advanced 1	64/51	1.25	
Advanced 2	47/23	2.04	
Total: Survey	436/376	1.16	

In both editions of the MAT, there appears to be a trend toward higher GR's at the upper levels of the test. The Science and Social Studies Tests in those upper levels include biographical material about famous people in those fields. As these achievement tests reflect current curriculum materials, which to a large extent still emphasize the contributions of men in these areas, it is inevitable that the tests contain an excess of male references. This may partially account for the higher proportion of male references in the upper levels of the complete Survey Batteries.

TABLE 2

Gender Ratios (GR) by Level for the 1970 MAT* & 1978 MAT Survey Tests

	1970 Edition		1978 Edition	
Test Level	MN + MP / FN + FP	GR	MN + MP / FN + FP	GR
Primary 1	51/59	.86	74/61	1.21
Primary 2	137/86	1.59	66/76	.87
Elementary	124/42	2.95	69/68	1.01
Intermediate	181/44	4.11	63/33	1.91
Advanced	198/51	3.88	64/51	1.25**
Median		2.95		1.21
Combined Across Levels	691/282	2.45	336/289	1.16

Table 3 shows the combined GR's at each level for the 1978 edition. These combined GR's were computed by adding the number of male nouns and pronouns and the number of female nouns and pronouns across the sets of Instructional Tests. The Instructional Tests do not show a trend similar to the Survey Tests of increased GR's at the upper level. This supports the hypothesis that the increased GR's of the Survey tests were a result of the Science and Social Studies items, since these two tests appear only in the Survey.

*1970 MAT data are from Faggen-Steckler, et al., 1974.
**Includes only the Advanced 1 level of the 1978 edition, the level corresponding to the 1970 Advanced level.

TABLE 3

Gender Ratios (GR) by Level of the 1978 MAT
Combined Across the Three Sets of Instructional Tests

Level	MN + MP / FN + FP	Combined GR
Primer	71/70	1.01
Primary 1	126/116	1.09
Primary 2	131/135	.97
Elementary	148/138	1.07
Intermediate	109/109	1.00
Advanced 1	123/77	1.60
Total	708/645	1.10

Since the 1978 MAT authors attempted to neutralize the language within the test, a neuter ratio (NR) of neuter nouns and pronouns to gender-specific nouns and pronouns was computed. These data are summarized in Table 4. The NR's ranged from .16 to 1.93 and tended to increase at the upper levels. The NR's were generally lowest in the Instructional Mathematics at the Primer through Elementary levels. At those levels, there is an abundance of items of the type: "Sally had 3 apples, she ate 1. How many does she have left?" While it is possible to rephrase those items using a neuter pronoun such as "I" or "we" in place of "Sally", to do so with all items might make the test monotonous. It should also be noted that neuter terms, especially neuter job titles and generic nouns are usually of greater reading difficulty (Beck, 1975). Since test item vocabulary is generally kept at or below grade-level, the rephrasing of items into neuter forms is restricted in the lower levels of the test.

TABLE 4

Neuter Ratio (NR): Neuter Nouns & Pronouns to Gender-Specific Nouns and Pronouns by Component and Level.

	$\frac{NN + NP}{(MN + MP) + (FN + FP)}$	NR	Median NR for Component
Instructional Tests			
Mathematics			
Primer	6/29	.21	
Primary 1	18/38	.47	
Primary 2	17/45	.38	.43
Elementary	10/46	.22	
Intermediate	30/24	1.25	
Advanced 1	29/18	1.61	
Reading			
Primer	27/44	.61	
Primary 1	55/115	.48	
Primary 2	67/95	.71	.73
Elementary	102/91	1.12	
Intermediate	128/88	1.45	
Advanced 1	76/100	.76	
Language			
Primer	29/68	.43	
Primary 1	70/89	.79	
Primary 2	87/126	.69	.76
Elementary	108/149	.72	
Intermediate	89/106	.84	
Advanced 1	85/82	1.04	
Survey Tests			
Preprimer	9/55	.16	
Primer	31/62	.50	
Primary 1	107/135	.79	
Primary 2	102/142	.72	.92
Elementary	143/137	1.04	
Intermediate	185/96	1.93	
Advanced 1	130/115	1.13	
Advanced 2	89/70	1.27	
Total Mathematics	110/200	.55	
Total Reading	455/533	.85	
Total Language	468/620	.75	
Total Survey	796/812	.98	
Components Combined	1829/2165	.84	

Table 5 lists each level across all components, with the thematic breakdown within each level. Traditional occupations, activities and emotions outnumber the nontraditional at almost every level for males and females alike. However, it is clear that females and males are also shown in non-stereotyped occupations and activities. The data indicate that more effort was made to balance the image of women than the image of men. For example, the ratio of females in traditional occupations to those in non-traditional ones is 26/28. The corresponding ratio for men is 90 to 7. The same pattern continues, though to a lesser degree, for the category of activities.

An encouraging result is that the ratio of female active roles to passive roles and the ratio of male active to passive roles are almost equal (2.53 and 2.41, respectively). Certainly not all females are portrayed as passive and dependent, nor are all males seen as strong and active. In fact, females and males are both shown in active roles more than twice as often as they are shown in passive roles.

The balance is somewhat reversed in the category of emotions. The majority of males and females display sex-stereotyped emotions. However, the range of emotions expressed by males was broader than that of women -- which is in itself non-stereotypic.

TABLE 5

Breakdowns by Sex and Thematic Category of Traditional and Non-Traditional Situations by Level Across Components

	Occupations				Activities			
	Female		Male		Female		Male	
	Trad	Non T	Trad	Non T	Trad	Non T	Trad	Non T
Primer	3	1	2	1	17	11	21	2
Primary 1	5	1	13	0	22	15	35	9
Primary 2	7	1	18	3	32	20	29	13
Elementary	3	3	16	0	33	20	39	20
Intermediate	5	9	22	1	18	10	21	2
Advanced 1	3	13	19	2	10	13	15	0
Totals	26	28	90	7	132	89	160	46

	Roles				Emotions			
	Female		Male		Female		Male	
Level	Active	Passive	Active	Passive	Trad	Non T	Trad	Non T
Primer	3	1	5	5	0	0	2	0
Primary 1	6	2	6	6	4	0	0	0
Primary 2	10	7	9	4	2	0	7	2
Elementary	5	3	9	1	2	2	1	0
Intermediate	6	0	2	1	1	2	4	1
Advanced 1	8	2	10	0	4	0	0	5
Totals	38	15	41	17	13	4	14	8

The thematic categories are somewhat difficult to interpret. ETS (1976) recommends that 25% or less of the items should be classifiable as "Traditional". Nevertheless, the issue can be framed in terms of whether the test ought to reflect society as it is (still with an overwhelming proportion of traditional models) or whether the test ought to show a more even distribution of non-traditional and traditional themes for women and men.

The gender balance within illustrations is summarized in Table 6. The ratios were computed for each level across components, and for each component across levels. With the exception of the Mathematics Instructional Tests, which contained very few pictures of people, the ratios of males to females are all very close to 1.00. This indicates an almost even split between female and male figures for the test as a whole and within each separate part.

TABLE 6

Ratio of Male to Female Figures in Illustrations
by Component and Level

	M/F	Ratio
Component (Across Levels)		
Instructional Mathematics	3/8	.38
Instructional Reading	44/44	1.00
Instructional Language	63/54	1.17
Survey	169/140	1.21
Total	279/246	1.13
Level (Across Components)*		
Primer	76/77	.99
Primary 1	118/107	1.10
Primary 2	55/40	1.38
Elementary	19/14	1.36
Total	268/238	1.13

The qualitative categories were rather inappropriate when applied to the drawings. Most drawings were not complete enough to permit an analysis of the figures' occupation, role, or emotion.

The ratios of number of females dressed traditionally to the number of females dressed non-traditionally ranged from .50 to 3.5. Most ratios (10 out of 14) were above 1.0. This could be accounted for by the fact that in today's society, dress is one of the simplest ways to discriminate between drawings of little girls and little boys.

The illustrations were coded with respect to the relative physical stature of males and females. Of the 45 drawings coded, 36 were of children and 9 were of adults. Table 7 provides a breakdown of these cases.

*The Intermediate, Advanced 1, and Advanced 2 levels contain no illustrations of people.

TABLE 7

Relative Physical Stature of Male and Female Illustrations:
Number of Children and Adults

	Children	Adults	Total
Male taller than Female	9	4	13
Female taller than Male	3	0	3
Female same as Male	24	5	29

In the majority of the illustrations, males and females of the same age were drawn as being of equal height (29 out of 45 cases.) In 13 out of 45 cases the males were taller. Among adults, such a portrayal may be true to life. However, 9 out of the 13 occurred in pictures of children. The major criticism in this area has been that boys are often shown taller than girls when, on the average, children of the same age are very similar in height regardless of sex. In only 3 cases (all among children) were females depicted as being taller than males.

The 1978 MAT authors and editors made an attempt to portray boys and girls and men and women in mixed-sex activities. Across all levels and components, there were 63 mixed-sex groups and 20 same-sex groups. Of the same-sex groups, 10 were composed of all females and 10 were all male. The attempt to illustrate mixed-sex group activities appears to have been successful.

Conclusions

It has often been assumed that biased item content affects pupil test performance. Whether or not this is true, and if true, exactly how content affects performance, are issues which have not been dealt with adequately in research literature to date. Without such relevant information, test publishers and critics have been relying on the concept of "balance" to minimize the possibility of bias.

There are also no data with respect to what an "appropriate" or ideal gender-balance would be and no objective way of determining how an achievement series ought to look. Test developers have assumed than an <u>even</u> balance would be optimal in reducing content bias.

Therefore, reducing the sex imbalance found in earlier versions of the MAT was one consideration in the construction of the 1978 MAT. The present analysis indicates there is a more even balance of male and female references, and generally, a less stereotyped portrayal of both sexes in the new version of the test.

The test is not evenly nor uniformly gender-balanced across all levels and components. However, given the problems of maintaining the appropriate grade-level vocabulary and the sex-bias/sex-stereotyping that exist in the current school curriculum, the 1978 MAT is certainly a significant step in the right direction.

References

Committee on Women in Words and Images. Dick and Jane as victims: Sex stereotyping in children's readers. Princeton, N.J.: Author, 1972.

Beck, M.D. Reducing "sex bias" in standardized achievement tests. Paper presented at the annual meeting of the American Personnel and Guidance Association, New York, March 1975.

Faggen-Steckler, J., McCarthy, K.A., & Tittle, C.K. A quantitative method for measuring sex "bias" in standardized tests. Journal of Educational Measurement, 1974, 11, 151-161.

Educational Testing Service. Guidelines for sex fairness in tests and testing programs. Princeton, N.J.: Author, 1976.

McGraw-Hill Book Company. Guidelines for equal treatment of the sexes in McGraw-Hill Book Company publications. New York: Author, 1974.

Saario, T.N., Jacklin, C.N., & Tittle, C.K. Sex role stereotyping in the public schools. Harvard Educational Review, 1973, 43, 386-416.

Science Research Associates, Inc. Fairness is educational materials: Exploring the issues. Chicago: Author, 1976.

Appendix A

Categories for Analysis of Gender-Balance

Written Materials

1. Occupations

   Female Traditional: Nurse, Teacher, Librarian, Secretary, etc.

   Male Traditional: Laborer, Professional, Principal, Boss, etc.

   Female Non-traditional: Professional, Laborer, Boss, Principal, etc.

   Male Non-traditional: Nurse, Teacher, Secretary, etc.

2. Activities

   Female Traditional: School, playing with dolls, onlookers, domestic chores

   Male Traditional: School sports, games, other physical activities, adventurer, etc.

   Female Non-Traditional: Sports, games, physical activity

   Male Non-Traditional: Domestic chores, child rearing etc.

3. Roles

   Active: Main character, problem solver, giving help/gift

   Passive: Secondary character, needing help, recipient of help/gift

4. Emotions

   Female Traditional: Fear, nurturance/tenderness, dependency, etc.

   Male Traditional: Aggression, courage, emotional strength, "strong silent type"

   Female, Male Non-Traditional: Cross-sex stereotypes

Appendix A (Continued)

Illustrations

1-4. Same as for Written Materials

5. Frequency of Male-Female Characters

6. Clothing

Female Traditional - dresses, hair ribbons, aprons, jewelry, etc.

Male Traditional - suits, trousers, uniforms etc.

Female Non-Traditonal - uniforms (except Nurses), pants, overalls, etc.

Male Non-Traditional - Nurse's uniform, apron, etc.

7. Physical Stature

Male/Female Same

Male Larger

Female Larger

within categories of adult or child

8. Grouping

Mixed Sex Group

Single Sex Group